(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57869 A2

(51) International Patent Classification[7]: G11B 20/00

(21) International Application Number: PCT/EP01/00979

(22) International Filing Date: 30 January 2001 (30.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/178,955 1 February 2000 (01.02.2000) US
09/537,815 28 March 2000 (28.03.2000) US

(71) Applicant: **KONINKLIJKE PHILIPS ELECTRONICS N.V.** [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor: **EPSTEIN, Michael, A.**; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: **FAESSEN, Louis, M., H.**; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PROTECTING CONTENT FROM ILLICIT REPRODUCTION BY PROOF OF EXISTENCE OF A COMPLETE DATA SET

(57) Abstract: A number of data items are selected for inclusion in a data set so as to discourage a transmission of the entire set over a limited bandwidth communications path, such as the Internet. Each portion of a data set is bound to the entirety of the data set so that portions of data sets that are independently distributed can be distinguished. In the case of audio recordings, for example, the data set includes an entire album, the individual songs on the album constituting portions of this data set. By binding each song to the album, a compliant player can be configured to refuse to render items in the absence of the complete data set. In this manner, the theft of a song requires a theft of the entire album. An uncompressed digital recording of an entire album consumes hundreds of megabytes of data, and the infeasibility or impracticality of downloading hundreds of megabytes of data is expected to provide sufficient discouragement for the theft of uncompressed content material. In a preferred embodiment, a watermark is created for each section of the data sent that contains an "entirety parameter" associated with the data set. The entirety parameter is a hash value that is based on a random number stored in the watermarks of each section. When presented for rendering, the entirety parameter is read, and the watermarks of a random selection of sections within the data set are compared to this entirety parameter to assure, with statistical certainty, that the entirety of the data set is present.

WO 01/57869 A2

Protecting content from illicit reproduction by proof of existence of a complete data set

This invention relates primarily to the field of consumer electronics, and in particular to the protection of copy-protected content material.

The illicit distribution of copyright material deprives the holder of the copyright legitimate royalties for this material, and could provide the supplier of this illicitly distributed material with gains that encourage continued illicit distributions. In light of the ease of information transfer provided by the Internet, content material that is intended to be copy-protected, such as artistic renderings or other material having limited distribution rights, are susceptible to wide-scale illicit distribution. The MP3 format for storing and transmitting compressed audio files has made the wide-scale distribution of audio recordings feasible, because a 30 or 40 megabyte digital audio recording of a song can be compressed into a 3 or 4 megabyte MP3 file. Using a typical 56 kbps dial-up connection to the Internet, this MP3 file can be downloaded to a user's computer in a few minutes. Thus, a malicious party could read songs from an original and legitimate CD, encode the songs into MP3 format, and place the MP3 encoded song on the Internet for wide-scale illegitimate distribution. Alternatively, the malicious party could provide a direct dial-in service for downloading the MP3 encoded song. The illicit copy of the MP3 encoded song can be subsequently rendered by software or hardware devices, or can be decompressed and stored onto a recordable CD for playback on a conventional CD player.

A number of schemes have been proposed for limiting the reproduction of copy-protected content material. The Secure Digital Music Initiative (SDMI) and others advocate the use of "digital watermarks" to identify authorized content material. European patent application EP 0981901 "Embedding auxiliary data in a signal" discloses a technique for watermarking electronic material, and is incorporated by reference herein. As in its paper watermark counterpart, a digital watermark is embedded in the content material so as to be detectable, but unobtrusive. An audio playback of a digital music recording containing a watermark, for example, will be substantially indistinguishable from a playback of the same recording without the watermark. A watermark detection device, however, is able to

distinguish these two recordings based on the presence or absence of the watermark. Because some content material may not be copy-protected and hence may not contain a watermark, the absence of a watermark cannot be used to distinguish legitimate from illegitimate material. On the contrary, the absence of a watermark is indicative of content material that can be legitimately copied freely.

Other copy protection schemes are also available. For example, European patent application EP0906700, "Method and system for transferring content information and supplemental information related thereto", presents a technique for the protection of copyright material via the use of a watermark "ticket" that controls the number of times the protected material may be rendered, and is incorporated by reference herein.

An accurate reproduction of watermarked material will cause the watermark to be reproduced in the copy of the watermarked material. An inaccurate, or lossy reproduction of watermarked material, however, may not provide a reproduction of the watermark in the lossy copy of the material. A number of protection schemes, including those of the SDMI, have taken advantage of this characteristic of lossy reproduction to distinguish legitimate material from illegitimate material, based on the presence or absence of an appropriate watermark. In the SDMI scenario, two types of watermarks are defined: "robust" watermarks, and "fragile" watermarks. A robust watermark is one that is expected to survive a lossy reproduction that is designed to retain a substantial portion of the original content material, such as an MP3 encoding of an audio recording. That is, if the reproduction retains sufficient information to allow a reasonable rendering of the original recording, the robust watermark will also be retained. A fragile watermark, on the other hand, is one that is expected to be corrupted by a lossy reproduction or other illicit tampering.

In the SDMI scheme, the presence of a robust watermark indicates that the content material is copy protected, and the absence or corruption of a corresponding fragile watermark when a robust watermark is present indicates that the copy protected material has been tampered with in some manner. An SDMI compliant device is configured to refuse to render watermarked material with a corrupted watermark, or with a detected robust watermark but an absent fragile watermark, except if the corruption or absence of the watermark is justified by an "SDMI-certified" process, such as an SDMI compression of copy protected material for use on a portable player. For ease of reference and understanding, the term "render" is used herein to include any processing or transferring of the content material, such as playing, recording, converting, validating, storing, loading, and the like. This scheme serves to limit the distribution of content material via MP3 or other compression

techniques, but does not affect the distribution of counterfeit unaltered (uncompressed) reproductions of content material. This limited protection is deemed commercially viable, because the cost and inconvenience of downloading an extremely large file to obtain a song will tend to discourage the theft of uncompressed content material.

It is an object of this invention to extend the protection of copy-protected material to include the protection of uncompressed content material.

This object and others are achieved by selecting a sufficient number of data items for inclusion in a data set so as to discourage a transmission of the entire set over a limited bandwidth communications path, such as the Internet. Each portion of a data set is bound to the entirety of the data set so that portions of data sets that are independently distributed can be distinguished. In the case of audio recordings, for example, the data set includes an entire album, the individual songs on the album constituting portions of this data set. By binding each song to the album, a compliant player can be configured to refuse to render items in the absence of the complete data set. In this manner, the theft of a song requires a theft of the entire album. An uncompressed digital recording of an entire album consumes hundreds of megabytes of data, and the infeasibility or impracticality of downloading hundreds of megabytes of data is expected to provide sufficient discouragement for the theft of uncompressed content material. In a preferred embodiment, a watermark is created for each section of the data sent that contains an "entirety parameter" associated with the data set. The entirety parameter is a hash value that is based on a random number stored in the watermarks of each section. When presented for rendering, the entirety parameter is read, and the watermarks of a random selection of sections within the data set are compared to this entirety parameter to assure, with statistical certainty, that the entirety of the data set is present.

The invention is explained in further detail, and by way of example, with reference to the accompanying drawings wherein:

Fig. 1 illustrates an example system for protecting copy-protected content material in accordance with this invention.

Fig. 2 illustrates an example data structure that facilitates a determination of the presence of an entirety of a data set in accordance with this invention.

Fig. 3 illustrates an example flow diagram of an encoder that creates a data set and accompanying parameters to facilitate a determination of the presence of an entirety of the data set in accordance with this invention.

Fig. 4 illustrates an example flow diagram of a decoder that renders a data item of a data set in dependence upon the presence of the entirety of the data set in accordance with this invention.

Throughout the drawings, the same reference numerals indicate similar or corresponding features or functions.

The theft of an item can be discouraged by making the theft more time consuming or inconvenient than the worth of the stolen item. For example, a bolted-down safe is often used to protect small valuables, because the effort required to steal the safe will typically exceed the gain that can be expected by stealing the safe.

The time required to download an MP3 encoding of an average song from the Internet, using a 56 kbs modem, is about 15 minutes, depending upon the network loading and other factors. Although it may impossible to define a specific "worth" to download time durations, it is believed that many people would be willing to incur a 15-minute download duration to receive a song of interest. On the other hand, the time required to download a non-compressed digital recording of an average song, using a 56 kbs modem, is about two hours, and it is believed that few people would be willing to incur a two hour download duration to receive a song of interest. Although a person may occasionally incur the two-hour download time to receive a song, the likelihood of two hour downloads becoming a prevalent means for stealing a song is expected to be minimal. For this reason, conventional protection schemes have been based on the need to identify compressed copies of protected material, using, for example, the combination of robust and fragile watermarks discussed above.

For ease of reference and understanding, the terms “lossless” and “uncompressed” are used synonymously herein. As will be evident to one of ordinary skill in the art, this invention is independent of whether the communicated information is compressed or uncompressed, and independent of whether the compression is lossy or lossless. The terms “compressed” and “uncompressed” are used herein because conventional solutions exist for the detection of lossy encodings such as MP3, and it is the degree of compression that is achieved by the lossy encoding of MP3 that has made wide-scale distribution of protected material feasible. As is known in the art, lossless compression

schemes exist. Lossless compression schemes, however, do not achieve the reduction in data that common lossy compressions achieve, and are not considered sufficiently "compressed" to be distinguished from "uncompressed" for the purposes of understanding this invention.

Also for ease of understanding, the invention is presented herein in the context of digitally recorded songs that are downloadable from the Internet. As will be evident to one of ordinary skill in the art, the invention is applicable to any recorded information that is expected to be transmitted via a limited bandwidth communications path. For example, the individual content material items may be data records in a larger database, rather than songs of an album.

The likelihood of a person illicitly downloading a song can be expected to be inversely proportional to the time required to effect the download. This invention is premised on the assumption that there is some threshold download duration above which the expected loss of revenue caused by illicit downloads is deemed acceptable. Experiments and surveys can be performed to determine a download duration that is sufficient to discourage the illicit download of a song, or, such a discouraging-duration can be estimated and will likely be in the order of ½ to 1 hour. That is, it is likely that "many" illicit downloads of a popular song will occur if the duration is less than a half hour, and "few" illicit downloads will occur if the duration is greater than an hour, even if the song is very popular.

As technology advances, and alternative communication schemes become available, the time required to download an uncompressed file can be expected to decrease. Using a DSL or cable connection to the Internet, for example, reduces the time required to communicate an uncompressed digital song to less than 5 minutes, depending on network loading and other factors. As noted above, most existing protection schemes are not able to distinguish lossless copies of digital data from the original copy. Therefore, in a high-speed data transfer environment, the likelihood of lost revenue due to illicit downloads of uncompressed digital songs can be expected to increase significantly.

In accordance with this invention, individual songs on a compact disc (CD), or other medium, are bound to the entire contents of the CD, and a compliant playback or recording device is configured to refuse to render an individual song in the absence of the entire contents of the CD. The time required to download an entire album on a CD in uncompressed digital form, even at DSL and cable modem speeds, can be expected to be greater than an hour, depending upon network loading and other factors. Thus, by requiring that the entire contents of the CD be present, at a download "cost" of over an hour, the

likelihood of a theft of a song via a wide-scale distribution on the Internet is substantially reduced.

Fig. 1 illustrates an example block diagram of a protection system 100 in accordance with this invention. The protection system 100 comprises an encoder 110 that encodes content material onto a medium 130, and a decoder 120 that renders the content material from the medium 130. The encoder 110 includes a selector 112 that selects content material from a source, and a recorder 114 that records this material onto the medium 130. The selector 112, for example, may be configured to select content information corresponding to songs that are being compiled into an album. The recorder 114 appropriately formats, encodes, and stores the information on the medium 130, using techniques common in the art.

In accordance with this invention, the encoder 110 includes a binder 116 that binds each item selected by the selector 112 to the entirety of the information that is recorded onto to the medium 130 by the recorder 114. In general terms, the information stored on the medium 130 constitute data items, the entirety of the information stored on the medium 130 forms a data set, and each data item is bound to the data set.

The decoder 120 in accordance with this invention comprises a renderer 122 and a gate 124 that is controlled by an entirety checker 126. The renderer 122 is configured to retrieve information from a medium reading device, such as a CD reader 132. As is common in the art, the renderer 122 retrieves the information by specifying a location index, and in response, the reader 132 provides the data located at the specified location index on the medium 130. Block reads of data at contiguous locations on the medium 130 are effected by specifying a location index and a block size.

The dotted lines of Fig. 1 illustrate an example song extractor 142 that extracts a song from the medium 130 and communicates it to an example CD imitator 144, representative of a possible illicit download of the song via the Internet. The CD imitator 144 represents, for example, a software program that provides information in response to a conventional CD-read command. Alternatively, the information received from the song extractor can be written to a CD medium, and provided to the conventional CD reader 132. As noted above, the song extractor 142 is likely to be used because the transmission of the entirety of the contents of the medium 130 is assumed to be discouraged by the purposeful large size of the contents of the medium 130.

In accordance with this invention, the entirety checker 126 is configured to obtain data from the medium 130, typically via the renderer 122, to determine whether the

entire data set is present. Any number of a variety of techniques, common in the art, can be used to verify the presence of an entirety of the data set. For example, a checksum corresponding to the data items in the data set can be used to verify that all the data items are present, by computing a checksum on the data items available to the renderer 112, and comparing this checksum to the checksum corresponding to the original entirety of the data set. This checksum can be digitally signed, communicated with the data set, and subsequently certified using a cryptographic key.

In the context of consumer devices such as audio CD playback devices, a checksum-based approach may not be feasible. Audio CD players include error-correcting and other decoding schemes that allow for a variance with each reading of the CD. Audio CD players do not, for example, necessarily start a playback of a song at precisely the same point. Similarly, if an error is detected while the CD is being read, a repetition of a prior section is often substituted for the erroneous section. A variance of a few bytes in the start of a song, or the repetition of a millisecond's worth of bytes, will not cause a noticeable audible difference, but the presence or absence of these bytes will have a significant effect on a checksum associated with the song.

Watermarks, and the corresponding watermark detection equipment, are configured to provide an accurate and repeatable reading of the watermark under a variety of circumstances. For example, a watermark is typically recorded at a substantially lower bit rate than the bit rate of the recorded audio signal, and redundant recordings of the watermark are used to further improve the likelihood that an accurate value is read as the watermark. As noted above, the robustness of a watermark can be varied, typically by varying the bit rate and redundancy of the recording of the watermark. Even a "fragile" watermark is typically configured to survive minor variances and anomalies that are common in the reading of information from a conventional consumer CD playback device. As used herein, the term watermark includes one or more watermark encodings; a watermark may include, for example, a fragile component and a robust component. Depending upon the watermark production method, these components may be embedded in the section independently, or as a common entity. For ease of understanding, the terms watermark, fragile watermark, and robust watermark are used herein independent of the method of collecting or segregating individual components during the watermark production process.

Fig. 2 illustrates an example data structure 200 for storing data items in a data set that facilitates a verification that the entirety of the original data set is present. A track 210 and section 220 structure is illustrated, consistent with the memory structure of conventional

CD and other storage media. In a preferred embodiment, the data set is self-referential: the data set contains one or more parameters that can be used to verify the presence of the other members of the data set. In the example data structure 200, a random value R(i) 234 is assigned to each section 220 of the data set. A hash H(R(i)) of each of these random values R(i) is stored on the medium preferably as "out of band" data (OBD) 240. This data 240, for example, may be stored within the Table of Contents of a typical CD, as "CD-ROM" data in a mixed audio-data CD, as a separate and unique data section, as a false song containing only data, and so on. A hash of a composite of the hashes H(H(R0), H(R1), ... H(Rn)) 240 is used as a check value CHK that identifies the entirety of the data set. The check value CHK 232 and the random value R(i) 234 form the watermark 230 that is associated with each section 220 of the data set. That is, in a preferred embodiment, the CD 130 of FIG. 1 is created with each recorded section 220 having a watermark 230 that includes an identifier CHK 232 of the entirety of the data set on the CD, and an identifying random number 234 of the section 220. Hash values are used because, in general, a hash computation is irreversible. The value used to produce the hash value cannot be determined, and the effects of a change of one or more of the items used to form the hash value also cannot be determined. (The term "cannot be determined" is used herein in the cryptographic sense: a determination of the value can be expected to consume more time and resources than is practical to pursue.) The section watermarks may be robust or fragile watermark. In a preferred embodiment, the check value CHK 232 is encoded as a robust watermark, to assure an identification of the material as protected material, and the random number R(i) 234 is encoded as a fragile watermark. As discussed above, a robust watermark is recorded at a lower bit rate or with more redundancy than a more fragile watermark. Alternatively stated, a fragile watermark consumes less resources than a robust watermark. Also as discussed above, a fragile watermark provides an indication of other forms of tampering, such as the compression of the protected data. The check value CHK may also be a portion, such as the lower m bits, of the hash of the composite. Although the security of a partial hash value is less than that of a complete hash value, the savings in resources to store this value may justify this reduction in security.

Fig. 3 illustrates an example flow diagram for an encoder 110 that creates a data set on a medium in accordance with this invention. At 310, a data item is selected for inclusion in the data set. This data item may be a song that is selected for inclusion in an album, a data record that is selected for inclusion in a database, and so on. The data item contains one or more data sections. For example, a song may be partitioned into a plurality of equal time-duration sections, each data record may form a single section, etc. A random

number R(i) is assigned to each section of the data item, at 320, and the size of the data item is added to an accumulated size of the entirety of the data set, at 330. In accordance with this invention, data items are added to the data set until the size of the data set is deemed large enough to discourage a subsequent transmission of the data set via a limited bandwidth communications channel. This "discouraging size" is a subjective value, and will depend upon the assumed available communications bandwidth, the loss incurred by the transmission, and so on. If the discouraging size has not been reached, at 335, another data item is selected for inclusion in the data set, via the branch back to block 310. Not illustrated in the example flow diagram of Fig. 3, other criteria may also be used to determine whether to add additional data items to the data set. For example, if the data items correspond to songs of an existing album collection, all of the songs will typically be added to the data set, regardless of whether the size of the data set has exceeded the determined discouraging size. If all of the songs of the album collection have been selected, and the discouraging size criterion has not yet been reached, other data items are selected to accumulate the required discouraging size. For example, data items comprising random data bits may be added to the data set to increase its size. These random bits will typically be stored as out of band data, CD-ROM data, and the like, to prevent it from being rendered as audible sounds by a conventional CD player. Alternatively, the data items may comprise other sample songs that are provided to encourage the sale of other albums, or images and video sections related to the recorded content material. Similarly, promotional material, such as Internet access subscription programs may also be included in the recorded information on the recorded medium. These and other means of adding size to a data set will be evident to one of ordinary skill in the art in view of this invention. In accordance with this invention, each of the selected data items are bound to the data set, such that a removal or alteration of any of the data items, including any random sections, promotional material, and the like that were added to increase the size of the data set, can be used to preclude the subsequent rendering of data items from this data set.

After the data items are selected to provide a sufficiently sized data set, a check value CHK is computed, based on a composite of the random numbers that were assigned to the sections of each data item, at 340. This composite may include, for example, a checksum corresponding to the random numbers, a checksum corresponding to a function, such as a hash function, of each random number, and so on. As discussed above, this CHK value is preferably a hash of the composite of the random values, or a portion of such a hash. At 350, a watermark is created for each section of the data set that includes this CHK value

and also includes the random number R(i) that is assigned to the section. As noted above, the CHK value is preferably encoded as a robust watermark, and the random number as a fragile watermark. Each section is recorded onto the medium with this composite watermark, at 360, and a hash of each section's random number is stored onto the medium, at 370, preferably as Out of Band Data (OBD). In this manner, the individual data items are bound to the entirety of the data set, via the CHK value, and the validity of this entirety value can be verified via the self-referential hash values of the random numbers that were used to create the CHK value. Other encoding schemes that bind the individual data items to the data set will be evident to one of ordinary skill in the art in view of this disclosure.

Fig. 4 illustrates an example flow diagram for a decoder 120 that is configured to render a selected data item, such as a selected song, in dependence upon the presence of the entirety of the data set associated with this data item. This flow diagram assumes that the encoding method of Fig. 3 had been used to create the original copy of the data item and data set. If another binding scheme is used, one of ordinary skill in the art will be able to appropriately modify the example flow diagram of Fig. 4 in view of this example embodiment. It is assumed that the flow of Fig. 4 is invoked upon a determination that the selected data item is a copy-protected item, via, for example, a detection of a copy-protect watermark or other identifying mechanism. At 410, the hashed values of the random numbers associated with each section of the data items in the data set are read from the medium, which may be a reading from the true copy (130 of Fig. 1) of the recorded content material, or from an illicit source (144 of Fig. 1). At 420, a hash is computed based on a composite of these hash values, using the same algorithm that is used at 340 of FIG. 3 to create the CHK value that is contained in the watermark of each section of a true copy of the content material. A random section of the data set is selected, at 430, and the check value of the watermark is read, as CHK', at 440. Alternatively, the first section that is selected for verification may be a section from within the data item selected for rendering, to immediately verify that the selected data item is part of the original data set. In the context of a recorded song, the data section typically corresponds to a fifteen second section of the song. If the computed check value CHK is not equal to the read check value CHK', at 445, indicating a modification to the collection of hashes of the random numbers, the decoder is configured to refuse to render the content material, at 480.

To verify that the data section has not been substantially modified, the random number assigned to the randomly selected section Sx is read, as R'(x), at 450. As discussed above, the random number is preferably stored as a fragile watermark, and a characteristic of

a fragile watermark is that a substantial modification to the data containing the watermark causes a corruption or destruction of the fragile watermark. A hash of the read random value H(R'(x)) is also computed at 450, and this hash H(R'(x)) is compared to the corresponding hash value H'(R(x)) that was read at block 410 and subsequently used to create the verification check value CHK. If these hash values do not match, at 455, the decoder is configured to refuse to render the content material, at 480. If these hash values match, other sections may be similarly tested, via the loop 465-430 until sufficient confidence is gained that the content material has not been substantially modified from the true copy. In a preferred embodiment, only one or two sections are tested, so as to minimize the delay introduced by this data-set-entirety verification procedure. When sufficient confidence is gained, at 465, that the entirety of the data set is present, the selected song is rendered, at 470. As will be evident to one of ordinary skill in the art, additional verification checks can subsequently be applied. Preferably, the watermark of each section of the selected song is verified as each section of the song is read, to verify that each section of the song is a valid member of the original data set, by checking that each CHK' value contained in each section corresponds to the verification check value CHK.

The foregoing merely illustrates the principles of the invention. It will thus be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the invention and are thus within its spirit and scope. For example, the examples presented above illustrate each part of the recorded material being part of the data set. In an alternative embodiment, select data items, or select parts of data items, may be used to form the data set, for efficiency purposes. For example, the tail end of songs may not be part of the "data set" as defined herein, because the watermark process may be based on a fixed block-size for each watermark, or each redundant copy of the watermark. If, for example, the watermark, or other parameter, requires ten seconds of a recording for a reliable embedding, the remainder of ((the song's length) modulo (10 seconds)) will be recorded on the medium, but not included in the "data set" whose entirety is being checked. In like manner, some promotional material may be included on the recorded medium, but purposely excluded from the data set, so that it may be freely copied and rendered elsewhere. Note also that the example flow diagrams are presented for ease of understanding, and the particular arrangement and sequence of steps are presented for illustration. For example, simple equalities are illustrated in the decision blocks for determining correspondence, whereas depending upon the particular techniques used to encode or decode the parameters, the assessment as to whether

the read item corresponds to a determined item can include a variety of intermediate processes. These processes may include, for example, a decryption of items based on particular keys, fuzzy logic or statistical testing to determine if two values are "close enough" to imply a correspondence, and the like. Variations such as these and others will be evident to one of ordinary skill in the art in view of this invention, and are included in the spirit and scope of the following claims.

CLAIMS:

1. A method for discouraging a theft of content material comprising:

collecting (310-335) a plurality of data items (210) comprising the content material to form a data set that is sized to be sufficiently large (335) so as to discourage a subsequent transmission of the data set via a limited bandwidth communication channel, and

binding (350-360) each data item of the plurality of data items (210) to the data set to facilitate a preclusion of processing of each data item in the absence of an entirety of the data set.

2. The method of claim 1, wherein the binding of the plurality of data items (210) includes

creating (350) one or more watermarks (230) associated with each data item

3. The method of claim 2, wherein the one or more watermarks (230) include

a robust watermark that is configured such that a removal of the robust watermark causes a corruption of the associated data item, and

a fragile watermark that is configured such that a modification of the associated data item causes a corruption of the fragile watermark.

4. The method of claim 1, wherein the binding of the plurality of data items (210) includes

creating (340) an entirety parameter (232) corresponding to the plurality of data items (210).

5. The method of claim 4, wherein the entirety parameter (232) is based on a hash function.

6. The method of claim 4, wherein

each data item includes one or more data sections (220),

each data section of the one or more data sections (220) having an associated section parameter (234), and

the entirety parameter (232) includes a hash of a composite of the section parameters (234) associated with the one or more data sections (220) of each data item.

7. The method of claim 6, wherein the section parameter (234) of the one or more data sections (220) includes a random number.

8. The method of claim 6, wherein the composite of the section parameters (234) includes a hash of each section parameter (234).

9. The method of claim 1, wherein the plurality of data items (210) includes a plurality of at least one of: digitally encoded audio content, and digitally encoded video content.

10. The method of claim 1, wherein the binding of each data item of the plurality of data items (210) includes:

assigning (320) a random number (234) to each section of each data item,

creating a section hash parameter (240) corresponding to a hash of the random number (234) of each section,

storing (370) the section hash parameter (240) of each section on a medium,

creating (340) an entirety parameter (232) corresponding to a composite of the section hash parameter (240) of each section,

creating (350) one or more watermarks (230) corresponding to each section, based on the entirety parameter (232) and the random number (234) of the section, and

storing (360) each section of each data item on the medium with its corresponding one or more watermarks (230).

11. The method of claim 10, wherein the one or more watermarks (230) include

a robust watermark that is configured such that a removal of the robust watermark causes a corruption of the associated data item, and

a fragile watermark that is configured such that a modification of the associated data item causes a corruption of the fragile watermark.

12. A method of encoding content material comprising:

encoding a plurality of data items (210) to form a self-referential data set that facilitates a determination of whether an entirety of the data set is present in a subsequent copy of the data set.

13. The method of claim 12, wherein the self-referential data set includes one or more hash values (240) corresponding to data items of the plurality of data items (210).

14. The method of claim 12, wherein the self-referential data set includes a hash value (232) corresponding to the plurality of data items (210).

15. The method of claim 12, wherein the plurality of data items (210) includes a plurality of at least one of: digitally encoded audio content, and digitally encoded video content.

16. The method of claim 12, wherein the self-referential data set includes an encoding of at least one hash value (232, 240) as a watermark (230).

17. The method of claim 16, wherein the watermark (230) includes:

a robust component of the watermark (230) that is configured such that a removal of the robust component causes a corruption of the associated data item, and

a fragile component of the watermark (230) that is configured such that a modification of the associated data item causes a corruption of the fragile component.

18. A method of decoding content material from a source comprising

determining (410-465) whether an entirety of the content material is present at the source, and

decoding (470) the content material in dependence upon whether the entirety of the content material is present at the source.

19. The method of claim 18, wherein determining whether the entirety of the content material is present includes:

reading (440) an entirety parameter (232) corresponding to the content material,

reading (410) a plurality of item parameters (240) corresponding to items of the content material,

determining (420) an entirety value from the plurality of item parameters (240), and

comparing the entirety parameter (232) to the entirety value.

20. The method of claim 19, wherein

the entirety parameter (232) includes a hash parameter corresponding to the content material, and

the determining of the entirety value includes

computing a hash value corresponding to the plurality of item parameters (240) to form the entirety value.

21. The method of claim 19, wherein determining whether the entirety of the content material is present further includes:

reading (450) an identifying parameter (234) from each of a selection of items comprising the content material, and

comparing (455) the identifying parameter (234) of each item of the selection of items to an identifier that is based on a corresponding item parameter of the plurality of item parameters (240).

22. The method of claim 21, wherein

the plurality of item parameters (240) includes a plurality of hash parameters,

each hash parameter of the plurality of hash parameters corresponding to a hash of the identifying parameter (234) of each item comprising the content material.

23. The method of claim 18, wherein determining whether the entirety of the content material is present includes:

reading (440) a first data set identifier (232) from a first item of the content material,

reading (440) a second data set identifier (232) from a second item of the content material, and

determining a correspondence between the first data set identifier and the second data set identifier.

24. The method of claim 18, wherein at least one of the entirety parameter (232) and the plurality of item parameters (240) is embedded as a watermark (230).

25. The method of claim 24, wherein the watermark (230) includes:

a robust component of the watermark (230) that is configured such that a removal of the robust component causes a corruption of the associated data item, and

a fragile component of the watermark (230) that is configured such that a modification of the associated data item causes a corruption of the fragile component.

26. The method of claim 18, wherein determining whether an entirety of the content material is present at the source includes:

reading (410) a plurality of section hash parameters (240) from the source,

computing (420) an entirety value based on the plurality of section hash parameters (240),

selecting (430) at least one select section (220) of the content material,

reading (440) a watermark value (232) of the at least one random section from the source, and

comparing (445) the entirety value to the watermark value (232).

27. The method of claim 26, wherein determining whether an entirety of the content material is present at the source further includes:

reading (450) a second watermark value (234) of the at least one select section (220) from the source,

hashing (450) the second watermark value (234) to produce a hashed watermark value and

comparing (455) the hashed watermark value to a section hash parameter of the plurality of section hash parameters (240) corresponding to the at least one random section.

28. A storage medium (130) that is configured to contain content material, the storage medium (130) comprising

a self-referential data structure (200) that is configured to contain a plurality of data items (210) of a data set corresponding to the content material, wherein

each data item includes one or more data sections (220),

each data section of the one or more data sections (220) having an associated section parameter (234),

wherein the entirety parameter (232) is based on a composite of the section parameters (234) of the plurality of data items (210), and facilitates a determination of whether an entirety of the data set is present in a subsequent copy of material obtained from the storage medium (130).

29. The storage medium (130) of claim 28, wherein the composite of the section parameters (234) is based on a hash of each section parameter (234).

30. The storage medium (130) of claim 28, wherein at least one of the entirety parameter (232) and the section parameter (234) of each data item are embedded in at least one watermark (230) that is associated with the data item.

31. The storage medium (130) of claim 30, wherein the at least one watermark (230) includes

a robust watermark that is configured such that a removal of the robust watermark causes a corruption of the associated data item, and

a fragile watermark that is configured such that a modification of the associated data item causes a corruption of the fragile watermark.

32. The storage medium (130) of claim 28, wherein the plurality of data items (210) includes a collection of at least one of: digitally encoded audio content, and digitally encoded video content.

33. An encoder (110) comprising:

a selector (112) that is configured to select a plurality of data items (210) to form a data set having a minimum size that discourages a communication of the data set via a limited bandwidth communications path,

a binder (116) that is configured to create one or more parameters (230, 240) corresponding to the plurality of data items (210) that facilitates a determination of whether an entirety of the data set is present at a decoder (120), and

a recorder (114) that is configured to combine the one or more parameters (230, 240) with the plurality of data items (210) to form a self-referential data set that is stored on a recorded medium.

34. The encoder (110) of claim 33, wherein the recorder (114) is configured to store the one or more parameters (230, 240) on the recorded medium as one or more watermarks (230) that are associated with one or more data items of the plurality of data items (210).

35. The encoder (110) of claim 33, wherein the one or more watermarks (230) include

a robust watermark that is configured such that a removal of the robust watermark causes a corruption of the associated data item, and

a fragile watermark that is configured such that a modification of the associated data item causes a corruption of the fragile watermark.

36. The encoder (110) of claim 33, wherein the one or more parameters (230, 240) include an entirety parameter (232) corresponding to the plurality of data items (210).

37. The encoder (110) of claim 33, wherein the one or more parameters (230, 240) include a plurality of section parameters (234, 240) corresponding to the plurality of data items (210).

38. The encoder (110) of claim 37, wherein the one or more parameters (230, 240) further include an entirety parameter (232) that is based on a composite of the plurality of section parameters (234).

39. The encoder (110) of claim 38, wherein the composite of the plurality of section parameters (234) includes a hash of each section parameter.

40. The encoder (110) of claim 37, wherein

each of the plurality of section parameters (234) includes a random number that is associated with the corresponding data item.

41. The encoder (110) of claim 33, wherein the plurality of data items (210) includes a plurality of at least one of: digitally encoded audio content, and digitally encoded video content.

42. A decoder (120) comprising:

a renderer (122) that is configured to receive data items (210) corresponding to a self-referential data set, and to produce therefrom a rendering corresponding to at least one of the data items (210), and

an entirety checker (126), operably coupled to the renderer (122), that is configured to preclude the rendering of the at least one of the data items (210) in dependence upon whether an entirety of the data set is present.

43. The decoder (120) of claim 42, wherein the entirety checker (126) is configured to:

read an entirety parameter (232) corresponding to the data set,

read a plurality of item parameters (240) corresponding to items of the data set,

determine an entirety value from the plurality of item parameters (240), and

compare the entirety parameter (232) to the entirety value.

44. The decoder (120) of claim 43, wherein the entirety checker (126) is further configured to:

read an identifying parameter (234) from each of a selection of items comprising the data set, and

compare the identifying parameter (234) of each item of the selection of items to an identifier that is based on a corresponding item parameter of the plurality of item parameters (240).

45. The decoder (120) of claim 44, wherein

the entirety parameter (232) includes a hash parameter corresponding to the data set, and

the entirety value includes a hash value corresponding to the plurality of item parameters (240).

46. The decoder (120) of claim 44, wherein the plurality of item parameters (240) includes a plurality of hash parameters,

each hash parameter of the plurality of hash parameters corresponding to a hash of a value associated with each item of the items of data set.

47. The decoder (120) of claim 44, wherein at least one of the entirety parameter (232) and the plurality of item parameters (240) are encoded as one or more watermarks (230) that are embedded in the data set.

48. The decoder (120) of claim 47, wherein the one or more watermarks (230) include

a robust watermark that is configured such that a removal of the robust watermark causes a corruption of the associated data item, and

a fragile watermark that is configured such that a modification of the associated data item causes a corruption of the fragile watermark.

49. The decoder (120) of claim 42, wherein the entirety checker (126) is configured to

read a first data set identifier (232) from a first item of the data items (210),

read a second data set identifier (232) from a second item of the data items (210), and

preclude the rendering of at least the second item in dependence upon a correspondence between the first data set identifier (232) and the second data set identifier (232).

50. The decoder (120) of claim 42, wherein the data items (210) a plurality of at least one of: digitally encoded audio content, and digitally encoded video content.

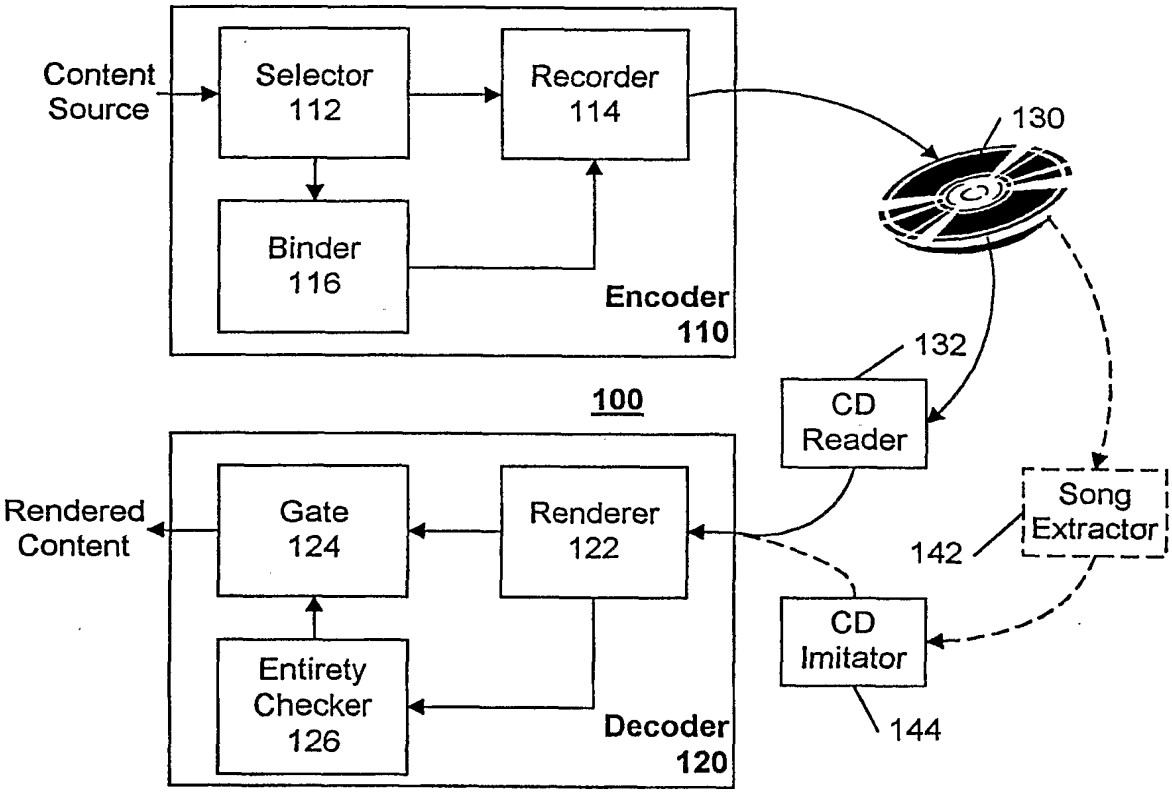
Content Source
Selector 112
Recorder 114
Binder 116
Encoder 110
130
132
CD Reader
100
Rendered Content
Gate 124
Renderer 122
Song Extractor
142
Entirety Checker 126
Decoder 120
CD Imitator
144

FIG. 1

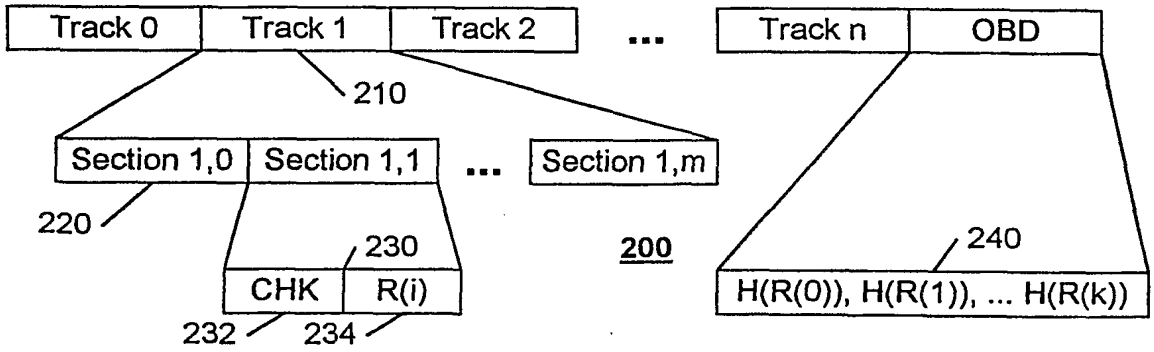
Track 0
Track 1
Track 2
...
Track n
OBD
210
Section 1,0
Section 1,1
...
Section 1,m
220
230
CHK
R(i)
232
234
200
240
H(R(0)), H(R(1)), ... H(R(k))

FIG. 2

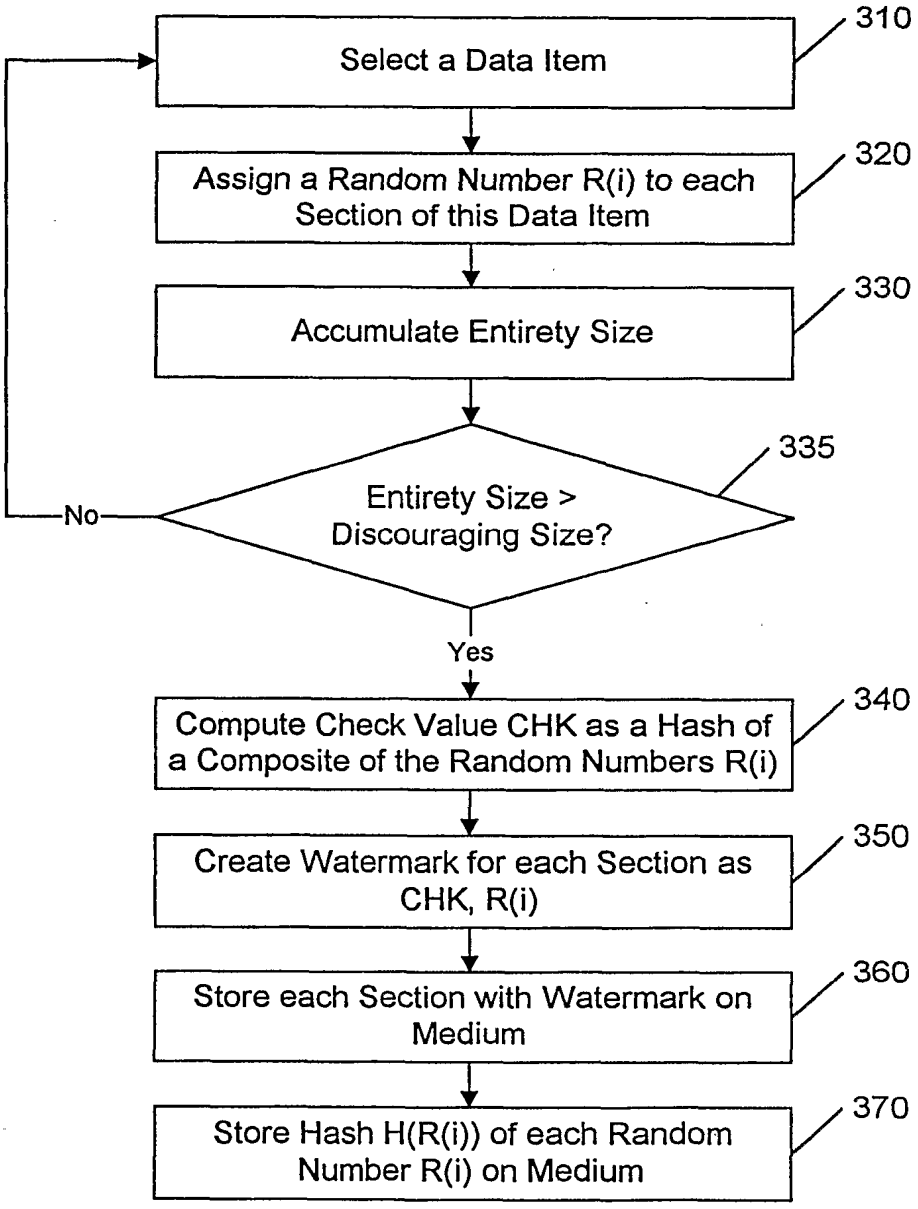
Select a Data Item
310
Assign a Random Number R(i) to each Section of this Data Item
320
Accumulate Entirety Size
330
Entirety Size > Discouraging Size?
335
No
Yes
Compute Check Value CHK as a Hash of a Composite of the Random Numbers R(i)
340
Create Watermark for each Section as CHK, R(i)
350
Store each Section with Watermark on Medium
360
Store Hash H(R(i)) of each Random Number R(i) on Medium
370

FIG. 3

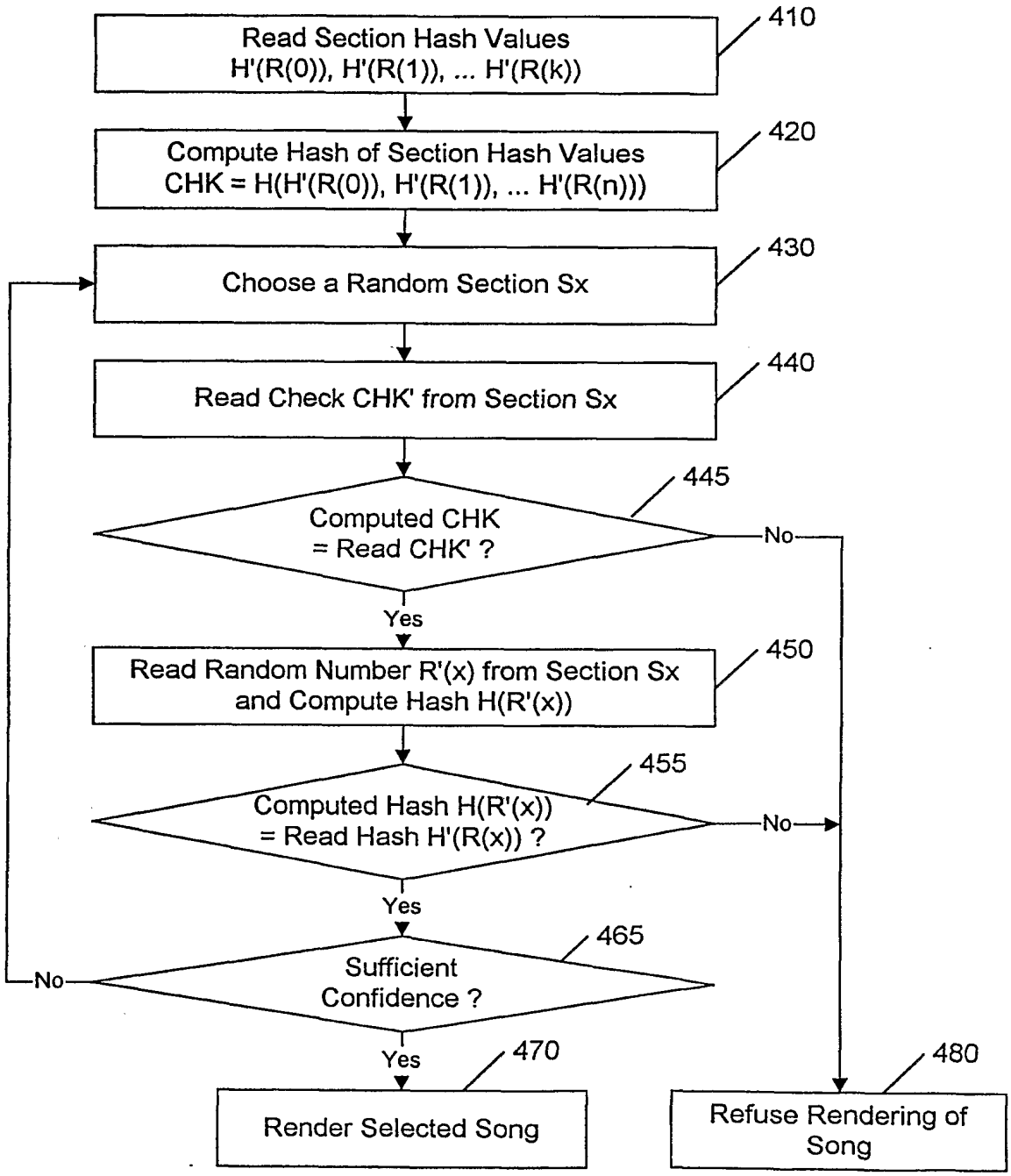
Read Section Hash Values
H'(R(0)), H'(R(1)), ... H'(R(k))
410
Compute Hash of Section Hash Values
CHK = H(H'(R(0)), H'(R(1)), ... H'(R(n)))
420
Choose a Random Section Sx
430
Read Check CHK' from Section Sx
440
Computed CHK
= Read CHK' ?
445
No
Yes
Read Random Number R'(x) from Section Sx
and Compute Hash H(R'(x))
450
Computed Hash H(R'(x))
= Read Hash H'(R(x)) ?
455
No
Yes
Sufficient
Confidence ?
465
No
Yes
Render Selected Song
470
Refuse Rendering of
Song
480

FIG. 4